DOCUMENT RESUME

ED 064 649 24 CG 007 323

AUTHOR Allen, R. R.
TITLE Terminal Report from the Concepts in Verbal Argument Project.
INSTITUTION Wisconsin Univ., Madison. Research and Development Center for Cognitive Learning.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
REPORT NO TP-35
BUREAU NO BR-5-0216
PUB DATE Dec 71
CONTRACT OEC-5-10-154
NOTE 24p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Cognitive Development; *Critical Thinking; *Learning Processes; Material Development; Measurement Instruments; Norms; *Research Methodology; Test Validity; *Thought Processes

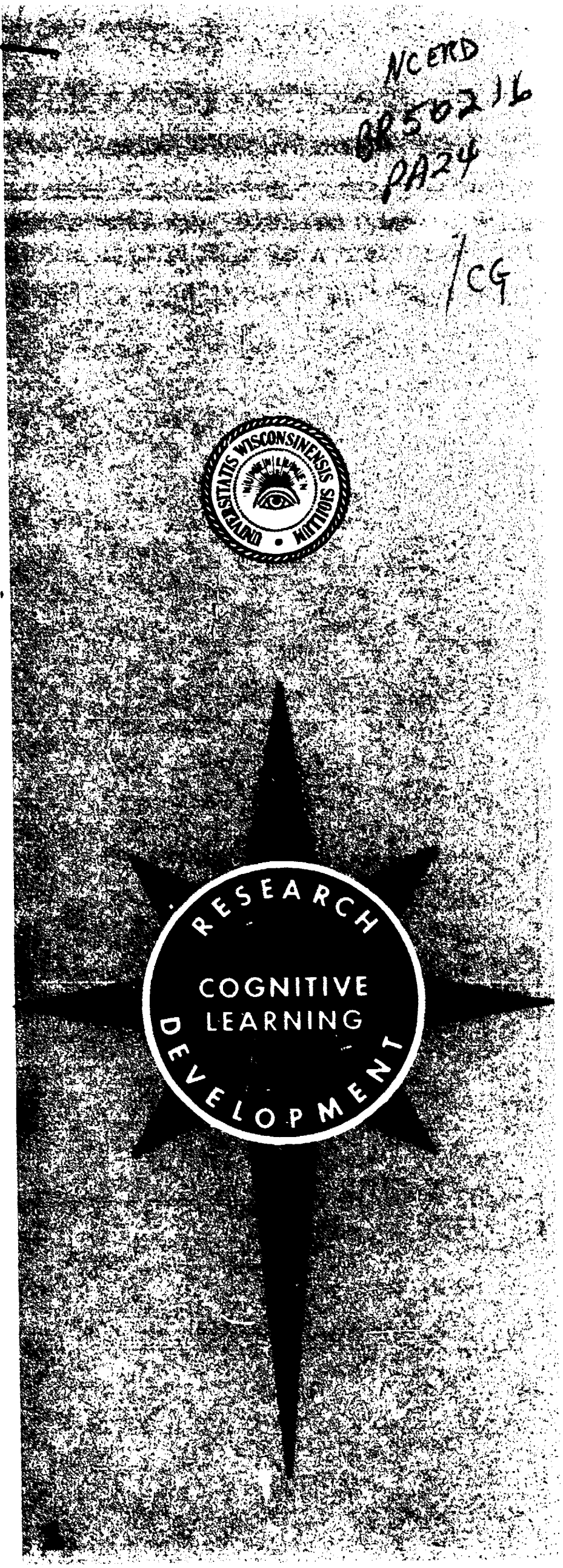
ABSTRACT

This terminal report summarizes nine phases of research and development activity of the Concepts in Verbal Argument Project: survey of the literature of critical thinking, identification of sequences of concepts and cognitive abilities, development of measuring instruments, factor analytic study of measuring instruments, normative study of student critical thinking abilities, development of instructional materials, field test of instructional materials, study of the effect of qualifiers on the acceptability of claims, and preparation of project reports. Special attention is given to the methodology and findings of studies related to test development and validation, establishment of norms for student critical thinking abilities, development and testing of programmed learning materials, and the effect of qualifiers in reason statements on the acceptability of claims. This report is intended to serve as a final general overview of the project. The reader wishing a comprehensive review of the project will wish to read the 12 other research and development documents produced by project personnel. (Author)

TERMINAL REPORT FROM THE CONCEPTS IN VERBAL ARGUMENT PROJECT

WISCONSIN RESEARCH AND DEVELOPMENT

CENTER FOR
COGNITIVE LEARNING



Theoretical Paper No. 35

TERMINAL REPORT FROM THE

CONCEPTS IN VERBAL ARGUMENT PROJECT

by
R. R. Allen, Principal Investigator

From the Project on Concepts in Verbal Argument

Wisconsin Research and Development
Center for Cognitive Learning
The University of Wisconsin
Madison, Wisconsin

December 1971

Published by the Wisconsin Research and Development Center for Cognitive Learning, supported in part as a research and development center by funds from the United States Office of Education, Department of Health, Education, and Welfare. The opinions expressed herein do not necessarily reflect the position or policy of the Office of Education and no official endorsement by the Office of Education should be inferred.

Center No. C-03 / Contract OE 5-10-154

Statement of Focus

The Wisconsin Research and Development Center for Cognitive Learning focuses on contributing to a better understanding of cognitive learning by children and youth and to the improvement of related educational practices. The strategy for research and development is comprehensive. It includes basic research to generate new knowledge about the conditions and processes of learning and about the processes of instruction, and the subsequent development of research-based instructional materials, many of which are designed for use by teachers and others for use by students. These materials are tested and refined in school settings. Throughout these operations behavioral scientists, curriculum experts, academic scholars, and school people interact, insuring that the results of Center activities are based soundly on knowledge of subject matter and cognitive learning and that they are applied to the improvement of educational practice.

This Theoretical Paper is from the Concepts in Verbal Argument Project in Program 2. General objectives of the Program are to establish rationale and strategy for developing instructional systems, to identify sequences of concepts and cognitive skills, to develop assessment procedures for those concepts and skills, to identify or develop instructional materials associated with the concepts and cognitive skills, and to generate new knowledge about instructional procedures. Contributing to these Program objectives, the staff of the project developed a semiprogrammed course in verbal argument and related tests for use at the high school level. The project staff prepared the materials on the basis of an outline of concepts and critical skills developed from an evaluation of everyday discourse.

Acknowledgements

As this project terminates, the Principal Investigator wishes to acknowledge his indebtedness to those who walked with him during his 5 years of research and development and to those who kept the path free of stones.

Special gratitude is owed to the co-investigators attached, at one time or another, to the project. Jerry Feezel and Fred Kauffeld, who joined the project in its second year, contributed richly to the tone and temper of the work in long afternoon chalk board sessions and late evening dialectics. Bill O'Brien, who served the project in its fourth year, contributed the insights of an experienced high school teacher to the creation of programmed learning materials. Dr. Robert Rott, who also served the project in its fourth year, contributed to test refinement and shepherded the massive normative study conducted in four separate Wisconsin school districts. Dr. Margaret Harris, who worked with the principal investigator in the final 2 years of the project, gave freely of her expertise in designing and implementing the various phases of testing.

In the spirit of the Wisconsin Research and Development Center for Cognitive Learning, a large number of other people contributed time and talent to project activities. The administrative, technical, and clerical staffs all provided sympathetic and competent support, often under pressures of time. The real impact of working within a community of researchers whose diverse competencies are brought to bear on problems of teaching and learning may never be fully measured. Finally, a great debt is owed to the director of the Center for his guidance and encouragement and for setting a standard of excellence against which all Center projects may judge themselves.

R. R. A.

Contents

List of Tables

Abstract

This terminal report summarizes nine phases of research and development activity of the Concepts in Verbal Argument Project: survey of the literature of critical thinking, identification of sequences of concepts and cognitive abilities, development of measuring instruments, factor analytic study of measuring instruments, normative study of student critical thinking abilities, development of instructional materials, field test of instructional materials, study of the effect of qualifiers on the acceptability of claims, and preparation of project reports. Special attention is given to the methodology and findings of studies related to test development and validation, establishment of norms for student critical thinking abilities, development and testing of programmed learning materials, and the effect of qualifiers in reason statements on the acceptability of claims.

This report is intended to serve as a final general overview of the project. The reader wishing a comprehensive review of the project will wish to read the 12 other research and development documents produced by project personnel.

I
Introduction

Statement of Project Focus

A number of subject matter fields in American secondary education have long professed to offer instruction relevant to student development of critical thinking abilities. Although the improvement of student critical thinking abilities has received widespread recognition as a worthy educational goal, few study groups and few teachers have been able to define well, even in a general way, what it means to think critically. As a consequence, direct instruction in critical thinking is usually absent from the schools.

In response to this condition, the Concepts in Verbal Argument Project at Wisconsin sought to identify and clarify the underlying conceptual structure of knowledge which enables student improvement in critical thinking abilities. Assuming the definitive stance that critical thinking is related to the assessment of claims and their justification against a system of rules appropriate to ordinary discourse, the investigators sought to lay out the structure of arguments established through testimony and arguments established through reasoning and to set forth relevant rules for the responsible appraisal of both types.

Having identified and clarified concepts in verbal argument, the investigators then sought to develop related measuring instruments for assessing student critical thinking abilities and programmed materials for teaching concepts in verbal argument at the secondary level. Two additional studies sought to determine the status of student critical thinking abilities and the effects of certain language variables on student assessment of verbal arguments.

Phases of the Project

The Concepts in Verbal Argument Project involved a number of interrelated research and development activities. Table 1 presents a project time schedule. Although the schedule may suggest that these phases are exclusive, it should be noted that each phase of the project informed, and in turn was informed by, the other phases. In the remainder of this introduction, each of the nine phases will be discussed briefly.

1. *Survey of the Literature of Critical Thinking.* A survey of relevant literature consumed the majority of the principal investigator's time during the first year of the project. In addition to surveying journals in education, educational research, speech, social studies education, and psychology, critical thinking tests and project reports from earlier long-term critical thinking studies were ordered and examined. During the first two quarters of the 1967-1968 project year, the survey of literature was updated prior to the preparation of a related paper (Allen & Rott, 1969).

2. *Identification of Sequences of Concepts and Cognitive Abilities.* The second major phase of the project involved the identification of concepts in verbal argument and their related cognitive (critical) skills. In formulating a taxonomy of such concepts and skills, the principal investigator and his two research assistants owed an intellectual debt to the fields of logic, rhetoric, argument, and semantics. In formulating the structure of knowledge related to argument through reasoning, the investigators modified a model based upon the logical construct formulated by Stephen Toulmin (1958). The Toulmin

Table 1. Concepts in Verbal Argument Project Time Schedule

Project Activity \ Academic Year	1964-1965				1965-1966				1966-1967				1967-1968				1968-1969		
	Sep.	Jan.	Apr.	June[a]	Sep.	Jan.	Apr.	June	Sep.	Jan.	Apr.	June	Sep.	Jan.	Apr.	June	Sep.	Jan.	Apr.
1. Survey of the Literature of Critical Thinking																			
2. Identification of Sequences of Concepts and Cognitive Abilities																			
3. Development of Measuring Instruments																			
4. Factor Analytic Study of Measuring Instruments																			
5. Normative Study of Student Critical Thinking Abilities																			
6. Development of Instructional Materials																			
7. Field Test of Instructional Materials																			
8. Study of the Effect of Qualifiers on the Acceptability of Claims																			
9. Preparation of Project Reports[b]																			

[a]Not funded under USOE Grant, Summer Session, 1965.

[b]Although project funding ended in June, 1969, the preparation of project reports continued well beyond that date.

approach to logical analysis was selected because it seemed better suited to both ordinary discourse and young minds than the traditional formal approach to logic. The first draft of the taxonomy was completed during the first two quarters of the 1965-1966 project year and was subsequently revised, following conferences with subject matter specialists, during the remainder of the 1966 calendar year. The product of this phase, which formed the base for later work in test and learning program development, was presented in the form of an Occasional Paper (Allen, Feezel, & Kauffeld, 1967).

3. *Development of Measuring Instruments.* From the project's inception, the principal investigator recognized the need for an appropriate testing instrument. Earlier critical thinking tests based on field-invariant logics usually neglected the concepts and skills related to assessing testimony and discerning the relevance of an objection. Tests based on the highly mechanical procedures for induction and deduction prescribed by type logics are particularly vulnerable to this criticism. Few ordinary arguments involve questions which can be resolved by direct observation and still fewer involve questions which can be fully analyzed against the tidy categories required by such systems. The Wisconsin Tests of Testimony and Reasoning Assessment (WISTTRA) were developed to assess the student's ability to evaluate adequacy of testimony and to recognize the structure that is present in ordinary arguments and raise pertinent objections based on the rules of inference appropriate to that structure. Work on the seven tests which comprise the battery was begun in February, 1966, and continued through April, 1968. During that period, the instrument went through four experimental editions in which its focus was narrowed from Grades 7-12 to Grades 10-12 and its items were analyzed and revised in order to improve the item characteristics and total test reliability. Portions of the battery were pretested on four occasions and a normative study was conducted with a fifth edition of the battery. The tests proper are contained in a Practical Paper (Allen, Feezel, & Kauffeld, 1969). A Technical Report presents a discussion of the test development and reliability estimates and item statistics for the fifth edition (Allen, Feezel, Kauffeld, & Harris, 1969).

4. *Factor Analytic Study of Measuring Instruments.* This phase of the project was begun during the summer of 1968 and was completed in June, 1969. The purpose of this study was to determine, using factor analytic procedures, the underlying abilities or dimensions measured by WISTTRA. Since WISTTRA was based on a schema for classifying concepts and critical abilities related to verbal argument, this phase may be viewed as a study of the construct validity of the earlier taxonomic effort (Allen, Feezel, & Kauffeld, 1967). A complete report of this phase is available as a Center Technical Report (Harris, 1969).

5. *Normative Study of Student Critical Thinking Abilities.* The normative phase of this project was accomplished largely during the 1967-1968 project year. Since a primary goal of the project involved making available learning materials in verbal argument for use by high school students, it was decided early in the project that data should be gathered regarding the critical abilities of the general target population. Such data were intended to provide a basis for determining the grade level at which instruction in these concepts would seem most appropriate. It was also intended that the normative data would guide the investigators in the preparation of the learning program by offering precise information regarding pre-instructional student skills. This phase of the project is more fully discussed in a Center Technical Report (Rott, Feezel, Allen, & Harris, 1969).

6. *Development of Instructional Materials.* The development of instructional materials in verbal argument for use by students was one of the primary goals of the project. Most earlier long-range critical thinking projects had terminated in a series of general recommendations and guidelines. Only one other project, still in process, shows promise of developing materials sufficiently complete for classroom use without burdensome demands on the teacher.

In the second and third quarters of the 1966-1967 project year the investigators familiarized themselves with a number of programming strategies and made several crucial decisions regarding the particular format to be used in the development of learning materials in verbal argument. The sequence and number of lessons were determined at that time. It was also decided, in consultation with programmed learning specialists, that the material involved would lend itself to the use of a semiprogrammed format. In this format, concepts were first introduced and clarified in textbook (paragraph) format before linear frames were presented as a means of enabling the student to internalize the concepts, and criterion measures were used to enable the student to demonstrate his ability to apply

the concept to a new critical instance. From June, 1967, to October, 1968, the actual creation of programmed lessons was in process. Each of the 17 lessons was developed in three drafts. The first draft involved the initial presentation of concepts in paragraph form. The second draft was complete in that linear frames and criterion measures were included. The final draft was a carefully edited and informed revision of the second draft. The three drafts did not occupy discrete time periods; i.e., final drafts for earlier lessons were completed before first drafts for later lessons were underway. Multiple copies of the learning program were produced by Center personnel in November, December, and January of 1968-1969 for limited dissemination and for field testing purposes (Allen, Kauffeld, & O'Brien, 1968a, b, c, d [Parts One, Two, Three, and Four]).

7. *Field Test of Instructional Materials* Consistent with Center policy, a field test of the instructional materials was conducted during the final year of the project. In December of 1968 and January of 1969 time was devoted to the planning of the field test and the development of necessary auxiliary materials. In February, March, and early April, 1969, the materials were used by more than 600 senior high school students in two Wisconsin school districts. Information regarding the field test is available as a Center Technical Report (Fischbach, Allen, & Quilling, in progress).

8. *Study of the Effect of Qualifiers on the Acceptability of Claims.* As the project continued to unfold, it became apparent to the investigators that a bias in favor of rationality in verbal argument had largely excluded consideration of certain semantic components of statements which may influence the listener's acceptance of verbal justification for claims. Although it was not feasible to examine the myriad of semantic factors which may influence the assessment of arguments, the investigators determined that the role of qualifiers in argumentative assessment was particularly worthy of investigation. Qualifiers are single words or word strings which are frequently used in statements to modify the strength of belief in the conjoined assertion. These have been discussed by theorists as weakening the commitment to an assertion (e.g.: probably, it's likely), or as functioning to strengthen commitment to an assertion (e.g.: I know, certainly). In either case, such words are imprecise representations of degrees of probability which are psychological rather than mathematical in nature. Since the very imprecision of qualifiers is a potential source of misunderstanding or misrepresentation in arguments, it was considered important to gain a greater understanding of the impact of these terms on receiver assessment of arguments. The design and conduct of this study occupied a portion of one research assistant's time during the 1967-1968 project year, and was subsequently used by that investigator as the basis for a dissertation submitted in partial fulfillment of the requirements of the Doctor of Philosophy (Communication Arts) degree at the University of Wisconsin. A Technical Report presents this research (Feezel, 1971).

9. *Preparation of Project Reports.* Although the principal investigator did not anticipate that the final phase of the project would involve the preparation of project reports, in fact a considerable portion of project time in the final year was given to such activity. Under the press of research and development deadlines, the project staff often found it necessary to move ahead to new activities without the luxury of reporting in print the products of earlier efforts. Three project reports required extension of time even after the project officially ended. The generous indulgence of those who have awaited these final project reports is appreciated.

II
Methods

General Statement Regarding Methodology

In a project of this scope and duration, diverse methodologies are employed. As the introduction to this paper suggests, the activities of the project covered a broad spectrum. At various moments, the investigators found it necessary to employ bibliographic, philosophical, and empirical methodologies. In the course of 5 years of study, the investigators were, in turn, bibliographers, theorists in verbal argument, test designers and evaluators, establishers of norms, authors of textual materials, conductors of field tests, etc. Rather than explaining each of these intellectual operations, this section will discuss the particular methods employed in accomplishing certain key project objectives.

Methodologies Related to Certain Project Objectives

Although the project may be viewed in terms of the various phases of project activity, project methodology probably is perceived best as it relates to important project objectives: the development and validation of testing instruments, establishing norms of student critical thinking abilities, the development and testing of programmed learning materials, and the study of the effect of qualifiers in reason statements on the acceptability of claims.

Test Development and Validation. As illustrated in Table 2, WISTTRA was constructed to measure cognitive skills related to certain fundamental concepts of verbal argument. The three tests of testimony were designed to measure the student's ability to detect instances which violate common internal and external tests of testimony. The reasoning tests were designed to measure the student's ability to recognize the essential components of an argument, to ask relevant questions about arguments, and to draw correct conclusions from arguments.

At two points in the development of the tests—before Pretest One and prior to the Normative Study—the battery was submitted to panels of argumentation experts trained in the conceptual basis of the instrument. On both occasions three-judge panels were used. Following a Q-sort technique, the judges were asked to place items in relevant categories or in a "cannot tell" category. Criteria for categorizing items included (where relevant) argument type, type of rule violated, statement type, and completeness of argument. Judge agreement ranged from 94.9 to 98.9% for the tests coded in the initial stages of development and from 85.4 to 98.4% for the tests used in the normative study. The decline in coder agreement is attributable to the fact that only items which achieved high coder agreement were used in drawing up the first edition of the tests, while the pool of items coded on the second occasion consisted of all items comprising the normative edition of the tests.

Hoyt analyses of variance reliability estimates were obtained for all of the tests. This is an internal consistency measure of reliability, and as such estimates consistency of performance on a relatively homogeneous power test.

During the development of the tests, items were continually revised to improve the instrument on the basis of item characteristic data obtained from the GITAP item analysis program (Baker, 1966; Baker & Martin, 1968). This program provides difficulty level, biserial correlation, X_{50}, and β statistics for each choice of each item. In addition, it gives descriptive statistics, the standard error of measurement, and the Hoyt reliability estimate for the total

Table 2. Relationship of WISTTRA to Concepts Identified

Appraising Testimony								Appraising Reasoning				
Internal Tests					External Tests			Data	Warrant	Reservation		Claim
TI 60 items					TII 20 items	TIII 40 items			RI 28 items	RII 28 items	RIII 28 items	RIV 28 items
Accept	Reject											
	Biased on topic	Not in position to observe	Incompetent to observe	Unqualified to judge	Consistency	Recency	Proximity	Always given—no test	Recognizing and selecting warrants	Recognizing statements which answer reservations	Selecting reservations	Selecting claims
20 items	10 items	10 items	10 items	10 items		20 items	20 items					
									Each of seven warrant types—sign, class, causal, alternative, parallel case, comparative, and warrant-supportive—is represented by four items in each reasoning test.			

test. Certain item characteristic criteria were used in selecting and refining items on the basis of the GITAP results. Items to be retained in a revised edition of the test had to meet the minimum requirement as given for each of the following criteria for the correct choice:

1. Preferably fall within a middle difficulty range as defined by Ebel (1965).
2. Have a biserial correlation $\geq .30$.
3. Have an X_{50} between +2.00 and -2.00.
4. Have a $\beta \geq .30$.

In addition each incorrect choice had to meet the following minimum requirements:

1. Have a reasonable minimum proportion of subjects respond to it.
2. Have a biserial correlation $< -.25$ and preferably $< -.30$.
3. Have an X_{50} lower than the X_{50} for the correct choice.
4. Have a $\beta < -.25$ and preferably $< -.30$.

These criteria were established in consultation with staff of the R & D Center and on the basis of reasonably standard rules of thumb for item evaluation.

An additional study, using factor analytic procedures, was designed to determine the underlying abilities or dimensions measured by WISTTRA. The WISTTRA battery was administered to approximately 3,000 students in Grades 7 through 12 in four Wisconsin school districts for the purpose of obtaining norms for the tests. The subjects for the factor analytic study consisted of 6 of the 12 groups from the normative study: boys and girls for Grades 8, 10, and 12. The total number of subjects within a single age and sex group studied ranged from 179 to 258. The number of subjects, by group, was: Grade 8 males, 200; Grade 8 females, 187; Grade 10 males, 223; Grade 10 females, 258; Grade 12 males, 179; and Grade 12 females, 240.

The treatment of the data consisted of two main procedures: reliability estimation and factor analysis. The data were analyzed separately for each grade and sex group.

Hoyt analysis of variance reliability estimates were obtained for each of the subtests for each sex and grade group studied. Means, standard deviations, and the intercorrelations of the 39 subtests were computed.

Three initial factor solutions were obtained: Alpha (Kaiser & Caffrey, 1965), Harris $R\text{-}S^2$ (Harris, 1962), and Unrestricted Maximum Likelihood Factor Analysis (UMLFA) (Jöreskog, 1967). A critical value of .05 was used to determine the number of factors for the UMLFA method. Each of these initial solutions was transformed by the normal varimax criterion (Kaiser, 1958) to give a derived orthogonal solution and by the Harris-Kaiser independent cluster method (Harris & Kaiser, 1964) to give a derived oblique solution.

The common factors from each of the six derived solutions were compared and the comparable common factors, those that are robust across solutions, were determined according to an interpretation strategy suggested by C. Harris (1967) and developed by M. Harris and C. Harris (1970).

Establishment of Norms for Student Critical Thinking Abilities. A normative study was conducted during the 1968 spring semester in the junior and senior high schools (Grades 7-12) of Clinton, Cedarburg, Reedsburg, and Owen-Withee, Wisconsin. More than 3,000 participating subjects were given the seven-test WISTTRA battery.

The mean and standard deviation were computed for each sex and grade group for each total test and for the subtests of Testimony I and Testimony III. The difference between the means of adjacent grades was found, by sex, for each of the total tests.

Intercorrelations of the seven tests in the WISTTRA battery were obtained. Included also were the intercorrelations of these seven tests with Testimony I as two subtests and with Testimony III as two subtests.

Intercorrelations of the seven tests in the WISTTRA battery with intelligence and reading scores were obtained for subjects from two of the schools, Cedarburg and Reedsburg.

Development and Testing of Programmed Learning Materials. The learning program was developed in 17 lessons organized in four parts. The content of the learning program is outlined in Table 3. Each of the lessons is written in such a way that concepts are presented first in paragraphs consisting of definitions and illustrations. The student then is asked to internalize the concept through linear frames which drill him on the concept to be learned. The linear frames are then followed by a branching frame, or criterion measure, which tests the student's application of the concept before permitting him to go on to the next concept.

The development of the learning program closely parallels the outline of concepts set forth in the earlier taxonomic work which was reviewed by three subject matter specialists. During the preparation of the early lessons,

Table 3. Contents of a Semiprogrammed Introduction to Verbal Argument

Part One: Argument in Perspective	
Lesson 1	Ordinary Uses of Language
Lesson 2	Language and Argument
Lesson 3	Language in Statements
Lesson 4	Statements as Claims
Part Two: Argument Through Testimony	
Lesson 5	Justifying Claims Through Testimony
Lesson 6	Internal Tests of Testimony
Lesson 7	External Tests of Testimony
Part Three: Argument Through Reasoning I	
Lesson 8	Justifying Claims Through Reasoning
Lesson 9	Sign Reasoning
Lesson 10	Individual to Class Reasoning
Lesson 11	Class to Individual Reasoning
Lesson 12	Reasoning from Alternatives
Part Four: Argument Through Reasoning II	
Lesson 13	Parallel Case Reasoning
Lesson 14	Causal Reasoning
Lesson 15	Comparative Reasoning
Lesson 16	Establishing Data and Warrants
Lesson 17	Qualifying Claims

Table 4. Number Correct on Pretest of Lessons 1-6 of the Learning Program

	Lesson											
	One		Two		Three		Four		Five		Six	
Student	32 Linear Frames	14-item Test	61 Linear Frames	21-item Test	36 Linear Frames	18-item Test	68 Linear Frames	26-item Test	16 Linear Frames	10-item Test	41 Linear Frames	16-item Test
1	32	14	58	21	36	16	68	26	16	10	41	16
2	32	14	61	21	36	18	67	23	16	10	40	16
3	32	14	58	21	35	16	67	23	15	10	41	16
4	31	14	60	21	36	17	65	26	15	9	40	16
5	31	13	59	21	36	17	67	26	16	10	41	16

the investigators consulted with a programmed learning consultant who offered detailed and careful criticism of the lessons. When the final edition of the first six lessons was completed, it was rated excellent by the programming specialist who recommended that his services were no longer needed. The first six lessons were then pretested with a sample of five sophomore students from Middleton High School. Table 4 provides data related to student accuracy in completing linear frames and lesson tests. Although the sample was extremely limited in size, the results tended to suggest that the investigators were not programming at too difficult a level.

In preparing all 17 lessons, consideration was given to readability and vocabulary level. The vocabulary was selected for a maximum level of ninth grade reading ability (Thorndike & Lorge, 1944). Upon completion of the 17

lessons and prior to the field test, a study of the readability of Lessons 1, 6, and 9 was conducted using the Dale-Chall (1948) Formula for predicting readability. Twelve 100-word samples were drawn from each of the three lessons and the average sentence length and the percentage of unfamiliar words were determined. When this information was applied to the Dale-Chall formula, the results showed the predicted readability of Lesson 1 to be 9th through 10th grade level, Lesson 6 to be 11th through 12th grade level, and Lesson 9 to be 11th through 12th grade level. The slightly higher difficulty of Lessons 6 and 9 is not alarming in that these lessons assume student familiarity with specialized terms presented in earlier lessons.

A field test was planned and conducted in the spring of 1969. The subjects were pupils in each of two randomly-chosen English classes from each of three tracks of pupils in Grades 10 and 12 in two Wisconsin high schools. Thus, approximately 50 pupils in each of six groups were tested in each school. The design for each school may be schematically represented as follows:

		Grade	
		10	12
Track:	High	50	50
	Average	50	50
	Low	50	50

The specific information obtained included:

1. Pre- and posttest performance and growth on cognitive tests measuring the acquisition and application of concepts presented in the instructional materials.
2. Postinstructional scores on eight-item lesson tests.
3. Attitudes of the various groups of pupils toward the materials, as measured by semantic differential scales.
4. Information on the length of time needed to complete each lesson in the program.
5. Error rate on the frames within each lesson for various groups of pupils.

From this information, judgments were made about the suitability of the materials for each of the groups tested.

To ascertain the separate effects of the instructional program, the pretesting, and motivation upon performance of pupils, an additional study was conducted using approximately 75 middle-track 11th grade pupils from one of the two high schools used in the field test. One class each was randomly assigned to the following treatments, where X indicates testing and T indicates use of the lessons:

Treatment	N			
1	25	X_1	T	X_2
2	25		T	X_2
3	25	X_1		X_2

From this design one can evaluate the effect of pretesting on sensitizing pupils to concepts to be learned, and thus infer the instructional value of a pretest used in conjunction with the materials. One also can separate the effects of the instruction and pretesting on final performance.

Data of the five types mentioned above were collected from Treatment Groups 1 and 2 (minus pretest data for Treatment Group 2); cognitive tests only were administered to Group 3. Data collected from pupils assigned to Treatment 1, or Treatments 1 and 2 if statistical analysis indicated no differences, could be used to provide information on the suitability of the program for middle-track 11th graders. (It is hypothesized that results of the first study will indicate that the lessons are difficult for average-track 10th graders, but suitable for average-track 12th graders, posing the question of suitability for the 11th graders.)

Study of the Effect of Qualifiers in Reason Statements on the Acceptability of Claims. The general purpose of this study was to examine qualified and unqualified argumentative reasons for their relative effects upon reader assessment of the strength of the conclusion. Qualifiers were selected to represent three degrees of probability (certainty, likelihood, possibility), three wording forms (adverb, impersonal pronoun adjective, personal thought), and three variations with respect to location (attached to data only, to warrant only, to both data and warrant).

Nine words and phrases were operationally defined as the variations in qualifier degree and form: Certainly, It is certain that, I know that, Probably, It is likely that, I believe that, Possibly, It is possible that, I suspect that. Although there are hundreds of qualifying words and word strings available for study, these qualifiers were selected for study because they are commonly used in everyday argument and because they have received particular attention in previous research and theory. A control condition with no qualifier (null) was

used for comparison with the independent variables.

The relative strength of acceptance of the claims was operationally defined by pairing each of the nine qualifiers (in argumentative context) with every other and the null to determine the stronger claim in each case (10 x 9/2 = 45 pairs). Scheffé's (1952) method for scaling responses to paired comparisons was modified to exclude the zero point. Thus *S*s responded on an eight-point scale for each pairing of conditions; the ends of the scale were assigned a value of four with the values descending to one for the two middle blanks.

One hundred and eleven 11th grade students in six social studies classes at Monroe (Wisconsin) Senior High School were randomly assigned, in approximately equal proportions, to three groups corresponding to the three levels of qualifier location.

For the main analysis, the positive and negative numbers for each of the nine qualifier conditions were summed yielding nine condition scores per *S* in each of the three location groups. All argument pairs containing the null condition were omitted from this main analysis, and each condition was summed across the remaining eight pairs, giving a possible score range of +32 to -32 for each *S*. The data were cast into a 3 by 3 by 3 analysis of variance model with repeated measures on degree and form, independent groups on location, and 37 *S*s per group as replicates. The BMDP08V analysis of variance program (Biomedical Computer Programs, U.C.L.A.) was used with the IBM 7040-7090 computer system at the University of Washington to analyze the data. An alternative analysis of variance for paired comparisons was calculated as outlined by Scheffé (1952) to enable comparison of the unqualified condition with each qualifier.

GPO 826-690-3

III
Findings

Test Reliability and Validity

The reliability estimates obtained for all of the tests for each age and sex group were sufficient for research purposes and the evaluation of group differences. In addition, for some of the tests (particularly for Grades 10-12), the reliability estimates were of a sufficient magnitude to allow for evaluation of differences among individuals.

The items, in general, exhibit the characteristics sought by the investigators. Many of the items fall within the middle difficulty range. Most items discriminate rather sharply, as indexed by high biserial correlations and βs. Most of the items which have low biserial correlations and βs are found in one of two tests, Testimony I or Testimony III, when total test score is the criterion measure. These low correlations may be indications that at least some items are measuring different abilities, and that subtests should perhaps be retained. Most of these same items have correlations and βs above .30 for the appropriate subtest when it is the criterion measure. As evidenced by the X_{50} item statistics, many more items are maximally discriminating among students of low and middle abilities than among students of high ability. Thus, these items are discriminating more clearly among less able students than they are among more able students. In general, the item statistics tend to increase in value from Grade 7 to Grade 12.

Although the final edition of the tests was designed primarily for Grades 10-12, there were indications that the tests might also yield useful information for Grades 7-9.

The major conclusion of the factor analytic study was that the tests based upon the taxonomy of concepts and abilities related to verbal argument as proposed by Allen, Feezel, and Kauffeld (1967) have construct validity at a particular level of specificity. The abilities underlying the assessment of verbal argument related to ordinary discourse seem to be the abilities to assess testimony in terms of internal (accept and reject) and external (consistency, recency, and proximity) tests of testimony, and the abilities to evaluate arguments developed through reasoning in terms of selecting the proper argument components of warrant, reservation, reservation answer, and claim. The type of warrant used in the argument did not seem to be of importance in terms of the underlying abilities represented by the comparable common factors. In teaching, however, one might still wish to make this distinction and use examples of reasoning for all of the warrant types.

All of the testimony subtests and most of the reasoning subtests were sufficiently reliable for research purposes.

The obtained factor structure, in terms of the comparable common factors, is quite similar for all groups studied but seems to be more clear for Grades 10 and 12 than it is for Grade 8.

It seems that, based upon the clarity of the comparable common factors, Grade 10 would be a good time to teach these concepts and abilities related to verbal argument as used in ordinary discourse.

The reasoning comparable common factors are fairly highly intercorrelated. The testimony comparable common factors are moderately correlated with the reasoning factors. The intercorrelations of the testimony factors tend to be low to moderate.

Student Abilities in the Evaluation of Verbal Argument

The mean scores tend to increase gradually from Grade 7 through Grade 12, with the stan-

dard deviation remaining fairly similar in most cases. With three exceptions, the magnitude of the differences between the means of adjacent grades is the greatest for any one test between Grades 9 and 10. For males the magnitude of the difference is the greatest between Grades 10 and 11 for Testimony II and Reasoning II. The one exception for females occurs for Testimony III; the difference is the greatest between Grades 7 and 8. These two exceptions for males and the fact that the differences between Grades 10 and 11 tend to be higher for males than for females are indications that male students may acquire the abilities tested a little later in life than do females. Looking at the total pattern of mean differences between adjacent grades, it seems that 10th grade may be a good time to teach these types of verbal argument skills. During this period students are acquiring many of these skills without instruction, and thus this may be the best time to supplement natural learning with instruction.

The intercorrelations of reading and intelligence scores were, in most cases, higher than the intercorrelations of reading or intelligence scores with scores on the testimony and reasoning tests. Correlations between intelligence and reading tend to be fairly high, while the correlations of the verbal argument tests with intelligence and reading tend to be low to moderate in magnitude. From this it would appear that the tests in the WISTTRA battery are measuring something different from the intelligence and reading tests.

Suitability of the Programmed Learning Materials

The main results pertain to the observed effectiveness of the lesson materials for their intended instructional purposes, their appropriateness for high school students, and characteristics of the field test itself, such as the method of administration which may have affected the effectiveness of the lessons during the field test.

Lesson Effectiveness. Analysis of gain scores, the difference between WISTTRA battery scores before and after the lessons were studied, suggests that gains were achieved on some, but not all, content areas by students at the 10th, 11th, and 12th grade levels. The results for the 10th and 12th grade students indicate that the lessons were most effective for upper-track students. The track groupings used by the schools were used to investigate lesson effectiveness in relation to ability or achievement level. The lessons did not appear to be effective for the lowest track in either grade or for middle-track sophomore students. Only middle-track juniors were studied. The gain scores of a randomly selected group which received the lessons were found to be higher on at least two batteries than another group which had not received instruction.

The greatest gains at all grade levels occurred on the WISTTRA battery subtests Testimony I and Reasoning II, while sophomores and seniors also showed gains on Testimony III and Reasoning I and III. All groups had lower, or no, gain on Reasoning IV and Testimony II, in order of postinstruction difficulty.

Lesson tests of eight items each were prepared for the field test to measure knowledge on each lesson immediately after its completion. A score of six correct was tentatively used as the criterion of minimum expected achievement for successful instruction. Mean number correct by lesson was computed by grade and track for the seniors and sophomores. For seniors the mean score was six or higher for seven lessons, and was lower than six for ten lessons. For sophomores the mean score was six or higher for only one lesson, and below six for sixteen lessons. Upper-track students, for both grades combined, achieved the successful instruction criterion in ten cases and failed in seven, while for the lowest-track students no means were six or higher and only for four lessons was the achievement of the middle-track students that high. Further analysis by grade and track indicated that upper- and middle-track seniors achieved the criterion on almost two-thirds of the lessons while upper-track sophomores did so on three-fourths of the lessons. Middle- and lower-track sophomores as well as lower-track seniors failed on all but one lesson.

Analysis of lesson frames completed by students while using the materials indicated that the number of correct responses followed about the same grade and track difference patterns as the other scores. Seniors had higher means than sophomores, but at both grade levels upper-track students had fewer errors. Lower-track sophomores achieved an error rate of less than five per cent on only three of the ten lessons analyzed, while middle-track sophomores did so on nine lessons and the upper-track students did so on all ten lessons. For seniors, the numbers by track are six for lower-track and ten for both middle- and upper-track students.

The overall implication of these results would seem to be that the effectiveness of the materials varies by content area and may be limited to the highest ability groups at lower grade levels in high school and to middle and higher ability groups among seniors.

Student and Teacher Reaction to the Field Test. The reaction of students to the materials may be indicative of the manner in which the field test was conducted or to characteristics of the materials or both. The presentation, which would normally take place intermittently over a semester or longer period, was compressed into a continuous session lasting about 6 weeks. The evaluations of the lessons by the senior and sophomore students were assessed after one-third, two-thirds, and all the lessons had been used. Since no zero point can be established, it is impossible to determine whether the students were favorable or unfavorable at any given time. However, it is clear that they became progressively less favorable or more unfavorable as the field test continued, as mean evaluative levels dropped on both the second and third assessments. The decrease occurred for all students regardless of grade or track although the decline was greater for seniors.

The teachers of the students studied were asked to inspect the materials and to complete a semantic differential evaluation form concerning the value, effectiveness, and appropriateness of the lessons for their students. Analysis of these responses suggests that the teacher reaction was generally quite favorable in all respects. Considered together, these evaluations suggest that the method by which the lessons were presented was detrimental to their effectiveness in the field test.

Effect of Qualifiers in Reason Statements on the Acceptability of Claims

The degree of the qualifier attached to reasons affects the strength of acceptance of the claim. The certainty degree terms led to significantly stronger acceptance than either the likelihood or possibility degree terms. The latter two degree groups did not differ significantly from each other.

The degree and the word form of the qualifier interact to determine the strength of acceptance of the claim. As degree increased, greater claim acceptance resulted for the personal thought form over the other forms; it seems the stronger the degree, the more weight is carried by an assertion of personal commitment. Comparisons among the nine qualifiers representing this interaction revealed these significant differences: (a) "I know" was stronger than the other certainty terms, which were in turn stronger than the other qualifiers; (b) "I believe" was stronger than "probably" with no other differences among likelihood terms; and (c) "probably" was not stronger than the possibility terms (the other two likelihood and the three certainty qualifiers were). Apparently there is an ambiguity about the degree of the term "probably."

There were no indications of interaction of qualifier degree with the reasoning component qualified (data, warrant, or both).

Comparison of the qualifiers with unqualified reasons in argument revealed that an unqualified statement effected significantly greater claim acceptance than any of the other qualifiers with the exception of "I know." This difference between no qualifier and the two certainty terms is the reverse of what some language analysts have suggested.

When comparing *Ss'* responses to qualifiers in argument with their responses to the qualifier words alone, the same significant main effect and interaction resulted but there were a few different findings for the individual mean comparisons. For qualifiers alone, all the likelihood terms were stronger than the possibility terms; this was not true for "probably" in argument contexts. Also "it is certain" was perceived as weaker and "it is likely" was perceived as stronger when in arguments than when alone.

It seems, generally, that although certain qualifying terms may be given relatively stable meanings by high school students, when such words are used in arguments the meanings may change somewhat. At any rate, qualifiers seem to represent a significant factor in adolescents' responses to arguments, though not necessarily in the manner suggested by language analysts.

References

Allen, R. R., Feezel, J. D., & Kauffeld, F. J. *A taxonomy of concepts and critical abilities related to the evaluation of verbal arguments*. Wisconsin Research and Development Center for Cognitive Learning, Occasional Paper No. 9, 1967.

Allen, R. R., Kauffeld, F. J., & O'Brien, W. R. *A semiprogrammed introduction to verbal argument: Part One, Argument in perspective*. Wisconsin Research and Development Center for Cognitive Learning, Instructional Materials, 1968a.

Allen, R. R., Kauffeld, F. J., & O'Brien, W. R. *A semiprogrammed introduction to verbal argument: Part Two, Argument through testimony*. Wisconsin Research and Development Center for Cognitive Learning, Instructional Materials, 1968b.

Allen, R. R., Kauffeld, F. J., & O'Brien, W. R. *A semiprogrammed introduction to verbal argument: Part Three, Argument through reasoning I*. Wisconsin Research and Development Center for Cognitive Learning, Instructional Materials, 1968c.

Allen, R. R., Kauffeld, F. J., & O'Brien, W. R. *A semiprogrammed introduction to verbal argument: Part Four, Argument through reasoning II*. Wisconsin Research and Development Center for Cognitive Learning, Instructional Materials, 1968d.

Allen, R. R., Feezel, J. D., & Kauffeld, F. J. *The Wisconsin tests of testimony and reasoning assessment (WISTTRA)*. Wisconsin Research and Development Center for Cognitive Learning, Practical Paper No. 6, 1969.

Allen, R. R., Feezel, J. D., Kauffeld, F. J., & Harris, M. L. *The development of the Wisconsin tests of testimony and reasoning assessment (WISTTRA)*. Wisconsin Research and Development Center, Technical Report No. 80, 1969.

Allen, R. R., & Rott, R. K. *The nature of critical thinking*. Wisconsin Research and Development Center for Cognitive Learning, Theoretical Paper No. 20, 1969.

Baker, F. B. *Test analysis package: A program for the CDC 1604-3600 computers*. Laboratory of Experimental Design, Department of Educational Psychology, University of Wisconsin, 1966.

Baker, F. B., & Martin, T. J. *FORTAP: A fortran test analysis package*. Laboratory of Experimental Design, Wisconsin Research and Development Center for Cognitive Learning, 1968.

Dale, E., & Chall, J. S. A formula for predicting readability: Instructions. *Educational Research Bulletin*, 1948, *27*, 37-54.

Ebel, R. L. *Measuring educational achievement*. Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1965.

Feezel, J. D. *The effect of verbal qualification of argument reasons upon acceptance of the derived claim*. Wisconsin Research and Development Center for Cognitive Learning, Technical Report No. 167, 1971.

Fischbach, T., Allen, R. R., & Quilling, M. R. *A field test of the semiprogrammed introduction to verbal argument*. Wisconsin Research and Development Center for Cognitive Learning, Technical Report, in progress.

Harris, C. W. Some Rao-Guttman relationships. *Psychometrika*, 1962, *27*, 247-263.

Harris, C. W. On factors and factor scores. *Psychometrika*, 1967, *32*, 363-379.

Harris, C. W., & Kaiser, H. F. Oblique factor analytic solutions by orthogonal transformations. *Psychometrika*, 1964, *29*, 347-362.

Harris, M. L. *A factor analytic study of the Wisconsin tests of testimony and reasoning assessment (WISTTRA).* Wisconsin Research and Development Center for Cognitive Learning, Technical Report No. 100, 1969.

Harris, M. L., & Harris, C. W. *A factor analytic interpretation strategy.* Wisconsin Research and Development Center for Cognitive Learning, Technical Report No. 115, 1970.

Jöreskog, K. G. Some contributions to maximum likelihood factor analysis. *Psychometrika,* 1967, *32,* 443-482.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika.* 1958, *23,* 187-200.

Kaiser, H. F., & Caffrey, J. Alpha factor analysis. *Psychometrika,* 1965. *30,* 1-14.

Rott, R. W., Feezel, J. D., Allen, R. R., & Harris, M. L. *Student abilities in the evaluation of verbal argument.* Wisconsin Research and Development Center for Cognitive Learning, Technical Report No. 106, 1969.

Scheffé, H. An analysis of variance for paired comparisons. *American Statistics Association Journal,* 1952, *47,* 381-400.

Thorndike, E. L. and Lorge, I. *The teacher's wordbook of 30,000 words.* Columbia: Teachers College Press, 1944.

Toulmin, S. *The uses of argument.* Cambridge, England: Cambridge University Press, 1958.

GPO 826-690-2